Clarifying DQO Terminology Usage to Support Modernization of Site Cleanup Practice
prepared by Deana Crumbling, USEPA Technology Innovation Office
May 30, 2001

Introduction

EPA's Technology Innovation Office (TIO) has been closely involved with developing and offering courses in both classroom and Internet web-based formats, with the goal of promoting the appropriate use of field analytical technologies and dynamic work plans since adoption of these tools and strategies could dramatically improve the cost-effectiveness of environmental restoration activities. TIO has found that a common language is vital if regulators, stakeholders, and practitioners are to negotiate, plan, and implement these projects to their mutual satisfaction.

Systematic planning is critical to the successful implementation of hazardous site characterization and cleanup projects. EPA's "DQO process" has been around for many years, and the "DQO" terminology is used extensively. Unfortunately, the terminology has been used in ambiguous or contradictory ways over the years. It is thus necessary to clarify the relationship between DQO-related terms as descriptively and concretely as possible. The discussion provided here has been reviewed by the primary DQO and data quality coordinators within the EPA Headquarters offices of the Office of Solid Waste, the Office of Emergency and Remedial Response, the Office of Environmental Information, and the Quality Staff, to ensure that the concepts presented are consistent with EPA's original intent for DQO terminology and with the direction that program needs are currently taking. Please direct any questions to Deana Crumbling, EPA Technology Innovation Office, (703)603-0643, crumbling.deana@epa.gov.

This paper is not intended to provide all-inclusive definitions that can be found elsewhere, nor does it attempt to provide an exhaustive coverage of each topic. It is intended to provide, as briefly yet unambiguously as possible, a basic conceptual understanding of DQO-related terms in a way that **facilitates systematic project planning in the context of site cleanups**. Descriptions for the terms/concepts appear first, followed by a discussion of the working relationships between the concepts. It is possible that other parties use terms other than these to communicate the same concepts. The actual terms used are less important than ensuring a clear understanding and application of the concepts, since these concepts are basic to the scientific validity of environmental decisions and to the data that support those decisions. A common conceptual understanding could help all within the hazardous waste community better communicate our goals and results, better plan and implement actual projects, and improve the cost-effectiveness and scientific defensibility of project decisions.

Descriptions for DQO-Related Terms

♣ Data Quality Objectives (DQO) Process–This is a systematic, iterative, and flexible planning process based on the scientific method. The DQO process was developed by EPA to provide a common structure and terminology to practitioners designing environmental data generation operations. It produces quantitative and/or qualitative statements (DQOs) that define project-specific decision goals, that will, in turn, guide the development of sampling and analysis

plans able to cost-effectively produce the "right kind of data." [More thorough discussions of DQOs and the DQO process can be found in various EPA Quality Assurance documents available through EPA's Quality Staff website: http://www.epa.gov/quality/qa_docs.html]

♣ Data Quality Objectives (DQOs) are qualitative and quantitative statements that translate non-technical project goals into technical project-specific decision goals. Project planners derive DQOs from the social, economic, and/or regulatory objectives of the environmental program that the data will ultimately be expected to support. **These goal-oriented statements establish the (technical) "bar" for overall decision quality or tolerable decision error in accordance with the (non-technical) objectives driving the project.** The DQOs for any particular project may or may not be highly specific in naming target elements, target media, action levels, along with the intended uses of the data. However, the less ambiguously the project DQOs are articulated, the more confidently and efficiently all parties will be able to work toward a successful project completion.

> *Example of a less detailed, quantitative DQO:* Determine with greater than 95% confidence that contaminated surface soil will not pose a human exposure hazard.
> *Example of a more detailed, quantitative DQO*: Determine to a 90% degree of statistical certainty whether or not the concentration of mercury in each bin of soil is less than 96 ppm.
> *Example of a detailed, qualitative DQO:* The turn-around time for lead (Pb) results generated by a field method shall support the real-time decision-making needs of the dynamic work plan to determine proper disposition of each bin of soil.

Even when expressed in technical terms, **DQOs should express "what"** (i.e., what decision) the data will ultimately support, **but should not specify "how"** that data will be generated (e.g., which analytical methods are to be used). Despite the name, "*Data Quality* Objectives," DQOs should be thought of as statements that express the *project objectives* (or decisions) that the *data will be expected to inform or support.* These project objectives serve to guide the eventual determination of the data quality that is needed to make good decisions, but **DQOs themselves should not attempt to directly define the specifics of that data quality**. Doing so short-circuits the systematic planning process, compromising the ability of project planners to optimize data collection designs to make projects more cost-effective (Step 7 of the DQO process). Some terms that more intuitively express the originally intended concept of "DQO" include "*Decision* Quality Objective" and "Project Quality Objective."

♣ Measurement Quality Objectives (MQOs) are project-specific analytical parameters derived from project-specific DQOs. MQOs include acceptance criteria for the data quality indicators (DQIs—see below) relevant to the project, such as sensitivity (e.g., what quantitation limit is desired), selectivity (i.e, what are the target analytes), accuracy (which includes bias and precision), completeness, etc. **MQOs establish the "bar" for analytical data performance parameters**, derived from the level of performance needed to achieve the project goals (as expressed in the DQOs).

However, project MQOs are **not** intended to be technology- or method-specific. As with DQOs, MQOs **specify "what"** the level of data performance should be, but **not "how"** that level of data

performance will be achieved. In other words, although MQOs provide the criteria for how good the data must be, MQOs do not specify exactly how the data must be produced, that is, MQOs do not specify what analytical method or technology is to be used. [In actual practice, during project planning, the planning team's analytical chemist will naturally be considering which specific technologies may be applicable even in the early stages of project planning. Evaluating and refining analytical options is a significant part of the iterative nature of systematic planning which seeks the most resource-effective work strategy that can achieve the stated project goals (i.e., the project DQOs). The project chemist should explore whether available innovative analytical technologies might achieve the project MQOs (i.e., the needed data quality), yet be able to do so in a way that is more resource-effective for the project than more traditional analytical options, perhaps because of lower per-sample costs or more rapid turnaround times.]

> *An MQO for one project might read:* "The overall precision of lead measurements taken on the soil in the bins must be less than 50% RPD when at least 10 samples are taken from each bin."
> *An MQO for a different project might read:* "The measurement method to be chosen must be able to detect the presence of compounds X, Y, and Z in groundwater at a quantitation limit of 10 μg/L with a recovery range of 80-120% and a precision of <20% RSD."

Because a large part of overall decision certainty is due to sampling considerations, derivation of the MQOs should be integral to developing a sampling design that balances the contributions of analytical uncertainties with sampling uncertainties to the overall decision uncertainty. [Discussions about the partitioning of decision uncertainty can be found in various statistical or sampling documents available at http://cluin.org/chartext_edu.htm#stats, including the 1990 *A Rationale for the Assessment of Errors in the Sampling of Soils*, and the 1989 *Soil Sampling Quality Assurance User's Guide*.]

♣ Data Quality Indicators (DQIs) are qualitative and quantitative measures of data quality "attributes." Quality attributes are the descriptors (i.e., the words) used to express various properties of analytical data. Thus, DQIs are the various measures of the individual data characteristics that collectively comprise the general, all-encompassing term "data quality."

Quality attributes (and the facets of data quality that they describe) include (but are not limited to) the following:

- selectivity/specificity (describes what analytes the technique can "see" and discriminate from other target analytes or from similar-behaving, but non-target, substances);
- sensitivity (describes the lowest concentration, or increment of concentration, that the technique can "see" or reliably quantitate);
- bias (describes how closely the technique produces results with a predictable deviation from the "true" value);
- precision (describes how much random error there is in the measurement process or how reproducible the technique is);
- completeness (describes whether valid data is produced for all the submitted samples, or just some fraction thereof); and

- comparability (describes whether two data sets can be considered to be equivalent with respect to a common goal).

The familiar "PARCCS parameters" consist of 6 primary DQIs that include measures of precision, accuracy (used in this context to denote bias), representativeness, comparability, completeness, and sensitivity. Precision, bias, and sensitivity describe properties that are readily measured quantitatively, so they are considered to be quantitative DQIs and they are controlled through the use of acceptance criteria within an analytical quality control (QC) program. Representativeness, comparability, and completeness are considered to be qualitative DQIs. Representativeness and comparability are critically important to the scientifically valid interpretation of analytical data, but estimating the degree of representativeness or comparability often requires the exercise of professional judgment in BOTH the science generating the data (e.g., analytical chemistry) and in the science involved in interpreting and using the data (e.g., designing a treatment system or modeling contaminant extent or migration).

There may be more than one DQI for a single data quality attribute. For example, the attribute of precision can be measured using mathematical formulas for relative percent difference (RPD), relative standard deviation (RSD), standard deviation (SD), variance (SD^2), and a variety of other calculations that can quantitatively express the degree of random fluctuation in a measurement process. The selection of a particular DQI to measure a specific data quality attribute (for example, selecting RPD to measure precision) is a matter of

- convention (what are people used to seeing);
- the characteristics of the analytical method (for example, does the method generate continuous or discontinuous data?);
- the data set being evaluated (for example, the formula for RPD cannot handle more than 2 values, whereas the formula for RSD can handle multiple values);
- and/or of the intended use for the data (which determines how extensively the quality of a data set must be documented).

The language of "data quality attributes" and "data quality indicators" provides data generators and data users with the conventions to determine (and the terminology to communicate) that a particular data set is of "known and documented quality" commensurate with intended data use. [A more thorough discussion of DQIs appears in EPA/QA G-5i, which will be publicly available as a peer-review draft in June 2001 at http://www.epa.gov/quality/qa_docs.html]

♣ A Demonstration of Proficiency shows that a particular operator or laboratory has the appropriate training and equipment to accurately perform a method. The demonstration may be done by using Performance Evaluation (PE) samples, or using known concentrations of analytes spiked into a clean matrix. The purpose of a demonstration of proficiency is to ensure that the performance of the operators and equipment is capable of producing data of known quality. [Proficiency demonstrations are discussed in Chapter 2 of the SW-846 Manual, available at http://www.epa.gov/epaoswer/hazwaste/test/chap2.pdf]

♣ A Demonstration of Applicability involves a laboratory study, pilot study, field trial, or other kind of activity that establishes the appropriateness and performance capability of a

particular method for a site-specific matrix and application. The purpose of a demonstration of method applicability is to ensure that a particular method or method modification can produce data of known quality, able to meet the project's decision goals, on the site- or project-specific samples to be tested.

♣ Systematic Planning is the process of clearly defining what the goals (i.e., primary decisions) of a project will be (including how much uncertainty will be tolerated); then deciding the types and amounts of data that will be needed to address the primary decisions. Optimization to achieve overall cost-effectiveness for a project may involve a mix of screening and definitive analytical methods whose roles are articulated in explicit decision logic. The DQO process is a systematic planning approach that EPA has articulated to aid *data collection and interpretation* activities. The DQO process does not address other aspects of project planning that are included under the broader term "systematic planning." Systematic planning also includes developing the work plans that will coordinate and guide site operations related to cleanup, worker safety, waste removal and disposal, public involvement and other activities needed to achieve project goals. Key to successful systematic planning in the involvement of sufficient technical expertise, generally provided through a multi-disciplinary team, that represents the scientific and engineering disciplines needed to adequately address all project issues. [EPA requirements for systematic planning can be found in various EPA policy statements, such as *Policy and Program Requirements for the Mandatory Agency-Wide Quality System* (EPA Order 5360.1 A2), available at http://www.epa.gov/quality/qs-docs/5360-1.pdf]

♣ Triad Approach–A strategy for cleaning up hazardous waste sites that relies on the integration of systematic planning, dynamic work plans, and rapid turnaround on-site measurements to reduce costs while maintaining or increasing the reliability and protectiveness of site decisions. [Discussion about the triad approach can be found in the paper, *Improving the Cost-Effectiveness of Hazardous Waste Site Characterization and Monitoring,* available at http://cluin.org/products/failsafe.htm]

The Relationships between Decision Goals, DQOs, MQOs, and QC Protocols

During project planning, there should a logical *conceptual* progression in the development of decision goals, DQOs, MQOs, and QC acceptance criteria. *In practice*, however, this will be an *iterative* process where various options for implementing a project are explored, dissected, and recombined, the feasibility and costs for various options are estimated and weighed, and then the most promising option is selected and fully developed into project work plans that will actually be implemented. As a project's planning documents (such as work plans, sampling and analysis plans, quality assurance project plans, health and safety plans) are developed and finalized, there should be a clear presentation of (and the reasoning behind):

- The general project decision goals;
- The more detailed, technical project goals (the DQOs), and the decision rules that will guide project decision-making;
- The goals for *data quality* (the MQOs);

- How sampling representativeness will be ensured, and how sampling uncertainty will be controlled;
- A list of the analytical technologies and methods selected;
- The QC protocols and criteria to be used with the methods to demonstrate that data of known quality are being generated; and
- A description of how data will be assessed and interpreted according to the decision rules.

At completion points in the project, reports summarizing the outcome of site investigations and remedial actions should also clearly reiterate the project goals and the means by which these goals would be achieved. This provides a benchmark against which actual achievements are assessed.

In the beginning of a project, high-level program managers often set the **broad, non-technical goals for projects**: For example, "Given a budget of $X, we want to clean up this lead contaminated soil in accordance with all environmental regulations and to the satisfaction of the residents in the neighborhood." The next question, of course, is "How do we do that?" So the next step is for the project manager or planning team to translate these broad, non-technical goals into more technically oriented goals that can address the specifics of considerations such as

- Regulations: What are the applicable environmental regulations? Are applicable action levels already in place in regulations, or do site-specific action levels need to be derived based on risk-drivers? If there is more than one possible regulatory action level, which one should be used?
- Confidence in the outcome: How certain do we need to be by the end of the project that we have indeed achieved regulatory compliance, and how will we demonstrate to regulatory agencies or stakeholders that the intended level of certainty has been achieved?
- Constraints: What are all the constraints that need to be accommodated (like seasonal weather, budget, property access, etc.)?

Making sure that no important details are left out of consideration is the purpose of a systematic planning process such as EPA's 7-step DQO process" [Detailed explanation of the DQO process as applied to hazardous waste sites can be found in the document, *Data Quality Objectives Process for Hazardous Waste Site Investigations (QA/G-4HW),* available through http://www.epa.gov/quality/qa_docs.html, and will not be duplicated here.] Statements that summarize the answers to these and other questions constitute "**the project DQOs.**" As noted earlier in this paper, the project DQOs consist of the unambiguous technical expressions of the overall project decision goals.

The next level of technical detail geared toward data collection involves translating the project DQOs into project **MQOs** [i.e., a general characterization of the kind of information (what parameters or analytes need to be measured, and what level of data quality for those parameters is needed) that will be needed to achieve the project DQOs]. Analytical data quality is most often only a very small part of the uncertainty that needs to be controlled in order to have sufficient

confidence in the actual project decisions. An honest examination of the "weak" links contributing to overall decision certainty may reveal that paying for expensive "definitive" analyses contributes nothing toward decreasing the overall uncertainty in the project decisions when there are larger uncertainties due to the limitations of sampling very heterogeneous media.

Sampling uncertainty is decreased when sampling density is increased. Composite sampling may sometimes be used to increase sampling density while lowering analytical costs. [Refer to EPA Observational Economy Series Volume 1: Composite Sampling, EPA/QA G-5S, and other statistical documents, all available from http://cluin.org/chartext_edu.htm#stats]. If composite sampling is undesirable, another way to cost-effectively increase sampling density is by using less expensive analytical methods (perhaps, using screening methods) in association with a well-planned QA/QC design and limited traditional analyses to provide data of known quality matched to the decision needs of the project. As long as the data quality can be demonstrated to be compatible with the project's decision rules, the confidence in the overall decision reliability that is gained by increasing the sampling density will not be lost by the use of a screening method. [For more details, see "Guidelines for Preparing SAPs Using Systematic Planning and PBMS" in the January/February 2001 *Environmental Testing & Analysis*. The article is available through http://cluin.org/chartext_edu.htm#planning].

When project planners wish to express desired decision confidence objectively and rigorously in terms of a statistical certainty level (that may have been specified in the project DQOs), statistical expertise is required to translate that goal into strategies that blend the number of samples, the expected variability in the matrix (i.e., heterogeneity), analytical data quality (e.g., precision, quantitation limits), the expected contaminant concentrations (i.e., how close are they expected to be to regulatory limits), sampling design (e.g., grab vs. composite), and costs into an interlocking whole. Since sampling design and analytical strategy interact to influence the statistical confidence in final decisions, the interaction between an analytical chemist, a sampling expert, and a statistician is key to selecting a final strategy that can achieve project goals accurately, yet cost-effectively. Software tools can assist technical experts to develop sampling and analysis designs. [See http://cluin.org/chartext_tech.htm#imp.]

The **statistician** is concerned with controlling the overall (or summed) variability (i.e., uncertainty) in the final data set, and with the interpretability of that final data set with respect to the decisions to be made. The statistician does this during project planning by addressing issues related to "sample support" (a concept that involves ensuring that the volume, shape, and orientation of extracted specimens are representative of the original matrix under investigation), by selecting a statistically valid sampling design, and by estimating how analytical variability could impact the overall variability. The **field sampling expert** is responsible for implementing the sampling design while controlling contributions to the sampling variability as actual sample locations are selected and as specimens are actually collected. The sampling expert does this by selecting and using sampling tools in ways that ensure that the sample support designated in the sampling plan is met in the field. The **analytical chemist** is responsible for controlling components of variability that stem from the analytical side (including aspects of sample preservation, storage, homogenization, subsampling, analyte extraction, concentration, and instrumental determinative analysis). The analytical chemist should select analytical methods

that can meet the analytical variability limits estimated by the statistician, and design an analytical QC program that defensibly establishes that those goals were met in the final data set.

Controlling the various sources of analytical and sampling uncertainties (assuming no clerical or data management errors) ensures that data of known quality are generated. Sometimes there may be only a single option available for a certain task, so the selection process is simple. Other times there may be more two or more options and cost/efficiency considerations can drive selection of the equipment and/or the design. It should be obvious that staff expertise (training and practical experience directly relevant to the techniques under consideration) is very important to project success.

The data characteristics that will **control analytical and sampling uncertainty** are articulated in the MQOs. Thus the MQOs specify "how good" the data must be *at a general level*. MQOs are thus contrasted with DQOs, which specify "how good" the decision must be. DQOs certainly are the ultimate drivers of how good the data must be, but DQOs themselves do not directly express data quality characteristics. Sometimes, as project planning progresses or as project implementation proceeds, it is discovered that a DQO is unattainable given the realities of the site conditions, the availability of suitable technology, and financial constraints. In collaboration with regulators and stakeholders, revision of the project DQOs may be required. For example, it may be discovered that current technology for a certain analyte is unable to provide the data needed to support risk decisions at a desired 10^{-6} cancer risk level. When a risk-based DQO is unachievable with current technology, an MQO known to be achievable with currently available technology may be substituted for the DQO. In other words, if it is clear that the ideal decision goal (the DQO) is unattainable, data quality goals (MQOs) based on the best available technology may be substituted for the ideal DQO until a time when newer technologies become available. It is important to note that the technology or method itself is NOT specified by the regulatory MQO. This allows the flexibility required for market incentives to encourage the development of technologies that can meet or exceed that same level of data quality more economically.

Although project MQOs are not meant to specify particular methods or technologies, they do serve to *guide* the selection of the technologies that can most cost-effectively meet the DQOs. As instrumentation is selected (based on factors such as the type of data needed, the turnaround time needed to support project activities, the expertise and infrastructure required to operate it, and costs), and as the analytical strategy for the project is perfected (perhaps including a "demonstration of method applicability"), analytical method SOPs and QC protocols are developed that are both method- and project-specific (i.e., tailoring an analytical method's performance to meet the specific data needs of the project). A QC protocol identifies the analytical parameter or DQI to be controlled, the limits within which results for that parameter are acceptable, and the corrective action procedures to be followed if those acceptance limits are exceeded. **QC acceptance criteria** should be very specific and should be designed such that if the QC acceptance criteria are consistently met, the project MQOs will be achieved, which means that the resulting data will be sufficient to meet the project DQOs and support the project decisions.

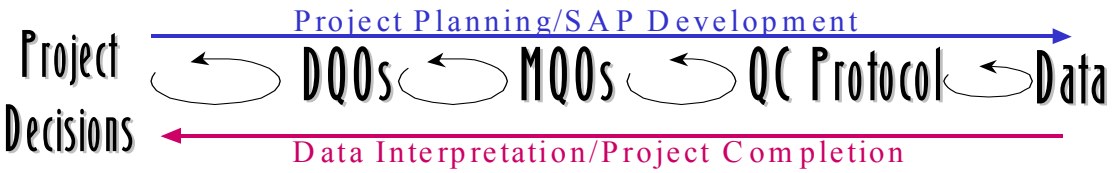
For example, an overall MQO for precision [for example, a statistically derived objective of less than <50% RPD between side-by-side (collocated) samples] may be partitioned into the primary components of variability that contribute to the overall variability. [Discussions about the partitioning of variability can be found in the "Rationale for the Assessment of Errors in the Sampling of Soils" document, available through http://cluin.org/chartext_edu.htm#stats webpage.] In the QC protocol, QC samples are used to monitor and document these measures of variability. The QC acceptance criteria are used to specify the maximum allowable variation in each component, and they might be expressed something like this:

- Analytical (instrumental) precision: "XRF instrument precision shall be determined using no fewer than 7 replicate analyses of a homogenized sample with a lead concentration near 400 ppm (the action level). The resulting RSD should be less than ±20%."
- Combined analytical and sample preparation precision: "Laboratory duplicates (prepared from a single sample with at least 150 ppm lead) should have RPDs less than ±35%."
- Combined analytical, sample preparation, and sample collection precision: "Field duplicates (collocated samples collected from a single location with at least 150 ppm lead, with each sample collected, prepared, and analyzed separately) should have RPDs less than ±50% (unless matrix heterogeneity is demonstrated to exceed the anticipated variability)."

The **figure** below serves to illustrate the conceptual progression that comprises the development of a design for generating data based on well-defined project goals. As stated earlier, while the conceptually this process is linear, in real-life, the development of a design is highly iterative, as portrayed by the circular arrows. The figure shows that the conceptual progression starts with the project-specific decision goals, and then moves "downhill" from broader, higher level goals to narrower, more technically detailed articulations of the data quality needs. Project decisions are translated into project-specific DQOs; then into project-specific MQOs; then into the technology/method selection and development of a method-specific QC protocol that blends the QA/QC needs of the technology with the project-specific QA/QC needs of the project. Finally, data are generated.

Then the process reverses. The actual raw data must then be assessed against the project MQOs to document that the quality of the data generated do indeed meet the decision-making needs of the project. The final step in the chain is interpreting the data into meaningful information (such as a statistical expression of a contaminant concentration as an average across an exposure unit) that is fed into the decision-making process (e.g., further action is or is not needed). If the "downhill" process has been conscientiously followed, there is a very strong likelihood that the "uphill" process of data assessment and interpretation will show that the data are of known and documented quality, and are fully adequate to support the project decisions.

The DQO Process



All data generation activities are derived from the

Project Decisions

so that data interpretation will lead back to and directly support the

Project Decisions.